# Problem Set 1

MACS 30010, Dr. Evans
Yangyang Dai

**Problem 1** Classify a model from a journal (5 points).
**Part (a, b)**

**Citation:**
Abdulkadiro?lu, Atila, Nikhil Agarwal, and Parag A. Pathak. 2017. "The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match." American Economic Review, 107(12): 3635-89.

**Part (c)**
In the paper, the difference in average student welfare between two matchings, $\mu$ and $\mu$', which are under coordinated and uncoordinated system, is given by

$$\bar{W}(\mu) - \bar{W}(\mu^{'}) = \frac{1}{|\zeta|}\sum_{i\in\zeta}(E[u_{i\mu(i)}|r_i] - E[u_{i\mu^{'}(i)}|r_i]) \tag{1}$$

where $\mu : \zeta$ denotes a matching,such that each student is assigned either to only one program or is unassigned, $r_i$ is the rank-order list, $\mu(i)$ is the student $i^{'}s$ assignment, $E[u...]$ represents the expectation of the utility for each student under coordinated mechanism, and  represents the expectation of the utility for each student under uncoordinated mechanism.

**Part (d)**
In the model, the average difference between student welfare $\bar{W}(\mu) - \bar{W}(\mu^{'})$ are endogenous, which is the output of the model. The $\zeta$, the utility of each student assigned by each program under coordinated system or uncoordinated system, $E[u...]$ and $E[u...]$ are exogenous, which are the inputs of the model.

**Part (e)**
This model is static as no time-dependent changes are introduced. It is linear as the endogenous variables are not interdependent. It is deterministic as no randomness is introduced.

**Part (f)**
Since the expectation of a student's utility under a specific program is conditioned on some other factors such as the demographic information of the program, or the students' home location. We could also add some factors that describe the individual traits, such as being independent or not, the willingness to move away from home and etc. into consideration when we categorize data.

**Problem 2** Make your own model. (5 points).

**Part (a,b,c)**
I will use a logistic regression model to estimate whether a person decides to get married or not.

$$Pr(Y_i = 1|X_{1i}, X_{2i}, X_{3I}, X_{4I}) = F(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4I} + \epsilon_i) \quad (2)$$

where

$$F(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

In my model, $Y$ is the outcome variable whose values 1 (get married) and 0 (not married).
$X_1$ is age, a continuous numerical variable; $X_2$ is a binary variable: whether has a stable relationship or not; $X_3$ is wealth, ta continuous numerical variable; $X_4$ is the degree of family support, on a 0-5 scale.
$\epsilon$ is the error term, representing haphazard events that can prompt a person to make the decision, such as unexpected family emergency. $\epsilon$ is drawn from a log-normal distribution $LN(0, \sigma^2)$.

**Part (d,e)**
The factors that affect my model the most would be the four i included in the part a-c, which are the age, whether has a stable relationship, wealth, the degree to which both families support this marriage.
I chose the variables described above based on my experience and study. Generally, those factors have more influence over people's decision to get married, as compared to the rest, for the following reasons:

1. age
    - People tend to have different attitudes towards marriage under different ages. In general, most people get married between 20 to 30.
2. wealth
    - People tend to decide to get married after they are financially stable, so they may be can afford raising children later on.
3. current relationship status
    - In general, people consider marriage when they think they are in a stable relationship.
4. family support
    - A common saying is that family's bless would make a marriage more happy. Whether parents' or friends' think positively of a person's potential marriage also could affect his or her decision to get married.

**Part (f).**
In order to do a preliminary test to examine whether my factors are significant in real life, i could use online anonymous survey such as the online platform Mturk and etc to collect sample data. After i collect the variables y(decision to marry), wealth, age, family support, relationship status, i could run the regression described above. If we want to see if our model matches up with real-world situation, we can simulate the data, using randomly generated $X_{1i}, X_{2i}, X_{3i}, X_{4i}$ and $\epsilon$ with similar mean and variance to the survey data, to predict $Y_i$. Then we compare the simluated $Y_i$ with $Y_i^{'}$ and see if the two groups are significantly different. If not, then we can say model is somewhat robust.